Computers & Education 135 (2019) 30–48

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The effect of cues for calibration on learners' self-regulated learning through changes in learners' learning behaviour and outcomes

Stijn Van Laer*, Jan Elen

Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, Bus 3773, B-3000, Leuven, Belgium

ARTICLE INFO

Keywords:
Functional validity feedback
Cognitive validity feedback
Self-regulated learning
Learning outcomes
Learning behaviour
Computer log files

ABSTRACT

Literature on blended learning emphasizes the importance of self-regulation for learning in blended learning environments and the role of learners' calibration. Although literature on calibration is clear on its importance for self-regulated learning, it provides inconclusive insight into the effect of support for calibration on learners' self-regulated learning. One under-investigated avenue might be learners' ability to act on the cues provided. In order to establish a more accurate picture of the effect of support for calibration on self-regulated learning, our study investigates whether providing cues for calibration affects learners' self-regulated learning, and whether this effect is different for learners with different metacognitive abilities. We investigate this effect by examining changes in learners' learning behaviour and outcomes. A pre-post design with one control and two experimental conditions was applied in a blended learning environment. Learners in the experimental conditions received either functional validity feedback (F condition) or functional and cognitive validity feedback (FC condition). Learners in the control condition did not receive any cues. Learners' behaviour was analysed using event sequence analysis. Learners' post-test learning scores were subjected to multivariate analysis of covariance, with condition and learners' metacognitive ability as independent variables. The results show a significant and unexpected impact of condition and learners' metacognitive abilities on learners' learning behaviour and outcomes. This manuscript discusses the unexpected results in terms of their theoretical and practical implications and provides recommendations for future research. We conclude that if cues for calibration are provided through functional and cognitive validity feedback, learners' calibration capabilities will increase. Yet, we hypothesize that for this to result in goal-directed self-regulated learning, learners might need to be guided in how to apply the cognitive and metacognitive strategies needed.

1. Introduction

Instructional interventions fostering self-regulated learning have been investigated widely in different educational settings (e.g., Arrastia-Chisholm, Torres, & Tackett, 2017; Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015), the actual effect of support for calibration on learners' self-regulated learning remains unclear (Panadero, Klug, & Järvelä, 2016). In general, literature investigating learners' calibration hypothesizes that learners are well calibrated if they perceive links between their learning behaviour, cues provided, information presented, and the task at hand, and when their perceptions reflect reality (Butler & Winne, 1995; Nelson, Narens, & Bower, 1990). In that case learners are equipped to effectively monitor their learning (DiFrancesca, Nietfeld, & Cao, 2016;

* Corresponding author.
E-mail address: stijn.vanlaer@kuleuven.be (S. Van Laer).

https://doi.org/10.1016/j.compedu.2019.02.016
Received 3 July 2018; Received in revised form 21 January 2019; Accepted 25 February 2019
Available online 01 March 2019
0360-1315/

Zimmerman, Schunk, & DiBenedetto, 2015). However, even when learners can calibrate external and internal feedback, they might not possess or be able to recall the cognitive and/or metacognitive strategies needed to act in a way that will produce increased learning outcomes (e.g., Pintrich, 2002; Veenman, 1993). This would result in sub-optimal self-regulated learning. Despite that literature on calibration is clear on its importance for self-regulated learning, it provides insufficient insight into how to support learners' calibration and so self-regulated learning (Stone, 2000; Yang, Potts, & Shanks, 2017). In order to establish a more accurate picture of the effect of cues for calibration on learners' self-regulated learning, this study investigates whether cues for calibration do actually affect self-regulated learning in a blended learning environment, and whether this effect is different for learners with different metacognitive abilities. We define blended learning as learning in an instructional context which is characterized by a deliberate combination of online and classroom-based interventions to instigate and support learning (Boelens, Van Laer, De Wever, & Elen, 2015, pp. 1–3) and we operationalize self-regulated learning as changes in learners' learning behaviour and outcomes. Investigating learning behaviour and outcomes provides insights on learners' self-regulated learning, as well as on the nature of cues' effects (Gašević, Dawson, & Siemens, 2015). In the next part of the introduction, we elaborate on blended learning and the conceptualization of self-regulated learning and present a theoretical basis for designing cues for calibration intended to evoke learners' self-regulated learning through monitoring. In the final part of the introduction, we discuss the relationship between self-regulated learning, learning behaviour, and learning outcomes.

1.1. Self-regulated learning

Like any learning, blended learning is a self-regulated process in which learners regulate their behaviour according to the instructional demands (Zimmerman & Schunk, 2001). By doing so, there proves to be a strong positive correlation between performance and self-regulated learning related variables (e.g., Daniela, 2015; Lin, Coburn, & Eisenberg, 2016). Self-regulated learning has been widely investigated and various theories have been proposed (see: Puustinen & Pulkkinen, 2001). Each of these theories describes self-regulated learning as cyclical, influenceable, and covert in nature. With regard to its cyclical nature, self-regulated learning roughly consists of three phases, namely (a) a forethought phase, (b) an enacting phase, and (c) an evaluation phase. Each theory also draws attention to the key role of internal (learner characteristics) and external factors (design of the learning environment) in the development of self-regulated learning (Zimmerman, 2013a). This cyclical and influenceable nature of self-regulated learning means that continuous measurements are needed to capture learners' self-regulated learning. Finally, regarding the covert nature of self-regulated learning, the metacognitive processes that occur in the different phases of self-regulated learning cannot be observed directly. If learners review their exercises after completing an online test, for example, it might be assumed that an evaluation or monitoring process must have preceded this overt cognitive reviewing activity. Given the dynamic nature of self-regulated learning, investigations would benefit from using continuous measurements and inferences that draw on learners learning' behaviour and learning outcomes (Panadero et al., 2016).

1.2. Monitoring, calibration, and self-regulated learning

Learners studying need to monitor their actions to ensure their actions lead to achieving the desired learning outcomes. Monitoring is the cognitive operation influencing whether action is taken or not (Muis, Winne, & Ranellucci, 2016; Winne & Jamieson-Noel, 2002). When learners interact with different tasks, information about changes in learning outcomes is monitored relative to learners' perceived changes. When discrepancies exceed an idiosyncratic threshold, self-regulating learners adjust their behaviour to eliminate the discrepancies (Winne & Hadwin, 1998). So for this adjustment to be effective, good calibration between perceptions of and actual changes in learning outcomes is needed. The better learners' calibration is, the more accurate monitoring will be (Stolp & Zabrucky, 2017). This accuracy is often referred to as judgement of learning (JOL) (Schraw, 2009). The information available to learners to calibrate and hence to monitor changes in learning outcomes has two main sources. Either changes in learning outcomes in reality (external feedback), or cognitive representations (internal feedback) by the learners of changes in learning outcomes (e.g., Ariel & Karpicke, 2018; Broadbent & Poon, 2015). For learners to be able to accurately calibrate, they have to process the external feedback along with the internal feedback (Winne & Hadwin, 1998). Learners have to compare (a) the internal feedback with the desired level of change in learning outcomes, (b) the external feedback with the desired level of change in learning outcomes, and (c) the internal feedback with the external feedback. Based on the results of this process, learners monitor their learning and select cognitive and metacognitive strategies (e.g., error correction strategies, revision activities, etc.) which may help them to proceed in the direction of the desired learning outcomes (Narciss, 2017, pp. 173–189).

1.3. Supporting learners' calibration for self-regulated learning

One reason for interest in learners' calibration is that learners use the result of comparisons of internal and external feedback to decide about how to monitor and self-regulate their learning. Thus, low levels of calibration and accurateness can undermine effective regulation (e.g., Dunlosky & Rawson, 2012; Thiede, Anderson, & Therriault, 2003). Despite its importance, a substantial body of literature suggests learners are generally not especially good at accurately judging themselves (Bol & Hacker, 2001; Klassen, 2002). And even when they are, extensive reviews by Dunlosky and Thiede (2013) and Alexander (2013) show that this does not mean that learners have any particular insight into this aspect of cognition.

To overcome this deficit and to provide support to learners' calibration attempts and evoke monitoring, cues for calibration are often suggested (e.g., Butler & Winne, 1995; de Bruin & van Merriënboer, 2017; Stone, 2000). Where prompts are extra support that is added in the form of hints or reminders to support a learner towards increased performance (Mayer, Sulzer-Azaroff, & Wallace, 2012, pp. 103–128), cues are elements in the environment, aimed to naturally remind the learner to perform a certain action and so support them towards increased performance (Hartley, Kieley, & Slabach, 1990).

Cues come in different formats, examples are: questions (e.g., Verpoorten, Westera, & Specht, 2017), information (e.g., Archer, 2010), and exercises (e.g., Slingerland, Weeldenburg, Raijmakers, Borghouts, & Vos, 2016). In contrast to prompts that from an self-regulated learning point of view might take over learners' self-regulated learning (e.g., Dinsmore, Alexander, & Loughlin, 2008), cues have a proven impact on learners' self-regulated learning as they potentially raise awareness with the learner about the actions needed and skills to be applied to perform goal-directed learning behaviour (Zimmerman, 2013b, pp. 10–45). Cues for calibration in this respect are hints aiming to induce the comparison of learners' own perception of competence with their true competence and so to support learners' calibration attempts and evoke monitoring (e.g., Azevedo & Hadwin, 2005; Zimmerman et al., 2015).

Literature on cues for calibration proposes two approaches, outcome feedback and cognitive feedback. Where outcome feedback (e.g., Delgadillo et al., 2017; Earley, Northcraft, Lee, & Lituchy, 1990; Paulson Gjerde, Padgett, & Skinner, 2017) is binary information describing whether or not results are correct, containing no additional information (e.g., about task, tools provided, or support offered) other than the state of the current learning outcomes (Butler & Winne, 1995), and providing minimal external support for learners about how to self-regulate their learning (Kluger & DeNisi, 1996). Cognitive feedback comes in two forms, namely functional validity feedback and cognitive validity feedback (e.g., Besser, 2016; Butler & Winne, 1995; Ernst & Steinhauser, 2017; Sedrakyan & Snoeck, 2016). Functional validity feedback, describes the relation between learners' estimate of change in learning outcomes and the actual change in learning outcomes (e.g., Bui & Loebbecke, 1996; Frysak, 2017; Popelka, 2015). For example, in an adaptive learning environment, learners might be asked to estimate their scores on a to-come test (in the form a JOL-cue). Then, after learners' estimates were compared to the actual score, functional validity feedback suggest to the learners, “You overestimated yourself, your score is 60% not 80%” (e.g., Mory, 2004). Cognitive validity feedback aims to evoke monitoring through the activation of learners' perceptions about the relationship between the different course components, information offered, cues provided, and potential change in learning outcomes (e.g., Chyung, 1996; Ellis, 2012, pp. 215–232). For example, in an adaptive learning environment, a learner who studies texts might be shown a cue: “You aren't using the advance organizer to guide your studying.” (Butler & Winne, 1995). Cognitive validity feedback conveys the information that directs learners' further actions based on their estimate and actual performance. Research investigating the use of functional and cognitive validity feedback does not lend uniform support to the effectiveness of these types of cues for calibration in practice.

1.4. Metacognitive skilfulness as influencing variable of calibration-cue use

Unarguably it is of vital importance to target cues for calibration so learners will eventually be directed to the right types of information. Nonetheless, cue-use research shows this is only one part of the challenge. It evidences that not all learners equally use and benefit from cues provided (e.g., Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011; Rashid & Asghar, 2016; Winne & Hadwin, 1998). Learners often do not utilize well-developed cues (Cleary, Callan, & Zimmerman, 2012). One of the possible explanations of sub-optimal cue use provided by the self-regulated learning literature is that learners might lack the skills needed or the ability to activate the targeted cognitive or metacognitive strategies to produce increased learning outcomes, this even when learners can calibrate external and internal feedback (e.g., Pintrich, 2002; Veenman, 1993). A way to investigate this issue and partly explain differences in the effect of the cues provided might be the investigation of learners' metacognitive abilities in relation to the instructional interventions provided (e.g., Ardasheva, Wang, Adesope, & Valentine, 2017). Metacognitive skilfulness or the ability to regulate and control cognitive and metacognitive strategies is a combination of general cognitive and metacognitive strategies. By mapping learners’ metacognitive skilfulness differences in the use of cues provided can be investigated (Veenman, Elshout, & Meijer, 1997). Doing so will provide us with a more fine-grained picture of the effect of cues for calibration.

1.5. Investigating learners’ self-regulated learning

Due to self-regulated learning's dynamic nature and its covert nature, continuous measurements and inferences through learners' learning behaviour and learning outcomes are needed to capture learners' self-regulated learning. Investigating both learners' learning behaviour and outcomes provides insights not only into learners' self-regulated learning, but also into the nature of cues' effects (e.g., Bannert, Molenaar, Azevedo, Järvelä, & Gašević, 2017).

1.5.1. Learning behaviour and self-regulated learning

As support grows for the conception of self-regulated learning as a continuous process, so does interest in online measures such as thinking-aloud protocols, eye-movement tracking and log-file registration, which account for the dynamic nature of learners' learning behaviour. Some of these methods (i.e., log file registration) involve no direct interaction with the learner, and this unobtrusiveness enables researchers to trace learning events in ecologically valid settings. The current study focusses on investigating differences in learners' learner behaviour through event sequence data analysis (Liu, Dev, Dontcheva, & Hoffman, 2016), so our focus lies on pattern mining defined as the identification of meaningful event sequences (patterns) (e.g., Azevedo et al., 2016; Bannert et al., 2015; Siadaty, Gašević, & Hatala, 2016). Pattern mining has two main concerns – the order of the events and the containment of sub-sequences. A sequential pattern implies the order of events has been preserved. A sub-sequence is a part of a larger sequence, which also appears elsewhere. Containment relates to the amount of support there is for a particular sub-sequence in the sample; in other words, the number (or percentage) of sub-sequences matching other learners' sub-sequences. A sub-sequence is considered frequent if it occurs in at least the threshold number of learners' sequences. Following sub-sequence identification, links between significant differences in subsequence occurrence and conditions either internal or external to the learner can be investigated in statistical trials.

1.5.2. Learning outcomes and self-regulated learning

Instruction may result in a variety of learning outcomes (e.g., Endedijk, Brekelmans, Verloop, Sleegers, & Vermunt, 2014). In this study, we focus on four main learning outcomes: (1) domain knowledge; (2) goal orientation; (3) academic self-concept; and (4) judgement of learning. We selected these learning outcomes as their relationship to self-regulated learning has already been investigated extensively. Domain knowledge, firstly, relates to learners' knowledge of the content involved in a particular task (Greene & Azevedo, 2007). Goal orientation was operationalized by Pintrich (2000) and Eccles and Wigfield (2002) in mastery and performance Objective, along with their approach and avoidance forms. Academic self-concept, the third outcome, can be defined as an individual's perception of self within academia (Elliot & Dweck, 2013) and incorporates the two distinct concepts of competence and effort. It measures the degree to which learners feel that academic subjects are easy and that they are good at them (competence), and the degree to which learners like or dislike going to school and studying different subjects (effort) (Liu & Wang, 2005). Finally, learners' judgement was operationalized in line with Schraw (2009), focussing on learners' precision of estimation of future performance compared to actual performance (Maki, Shields, Wheeler, & Zacchilli, 2005). When learners are able to accurately estimate their performance, they are more likely to take appropriate action and so regulate their learning (Butler & Winne, 1995). Learners who have a poor judgement of learning tend to make ineffective, suboptimal learning choices (Segedy, 2014).

1.6. Problem statement and hypotheses

While the literature emphasizes the importance of self-regulation for learning in blended learning environments on the one hand and the role of learners' monitoring through calibration for self-regulated learning on the other, evidence is inconclusive on the use of cues for calibration and their effect on self-regulated learning. Given this inconclusiveness, guidelines for interventions are difficult to outline. Hence, new approaches are needed to better understand the underlying mechanisms that may help to understand the inconclusive results. To get more profound insights in the effect of cues for calibration, this study investigates whether cues for calibration in blended learning environments foster self-regulated learning through changes in learners' learning behaviour and outcomes, and if this effect is different for learners with different metacognitive abilities. This leads us to four hypotheses:

Hypothesis 1. “Cues for calibration affect learners' learning behaviour.”

Hypothesis 2. “Cues for calibration affect learners' learning behaviour differently for learners with different levels of metacognitive skilfulness.”

Hypothesis 3. “Cues for calibration positively affect learners' learning outcomes.”

Hypothesis 4. “Cues for calibration positively affect learners' learning outcomes most when learners have high levels of metacognitive skilfulness.”

2. Method

2.1. Participants

The participants in this study were 151 learners taking a course on instructional psychology and technology as part of a Bachelor's degree in Educational Sciences from a large Belgian university. There were 134 women (88.74%) and 17 men (11.26%), which is a representative sample of the entire student population within the Faculty of Psychology and Educational Sciences. The learners were between 19 and 58 years of age ($M = 21.87$, $SD = 6.84$). They were familiar with the domain of instructional psychology and technology to some extent, but before the experiment they had not acquired insight in the texts of Anderson (2005) and Mayer (2004), which were the subject of the study task in the experiment. The subject matter was expected to be entirely new to them. This was controlled for in a prior domain knowledge test. None of the participants was able to achieve the maximum score on the test's questions, the average score was 4.5/10. It was concluded that the learners could be divided over the experimental groups at random. All learners voluntarily participated in the study, few (2) were excluded from the behavioural trace analysis because of incomplete data.

2.2. Content and module description

2.2.1. Content

In the course ‘instructional psychology and technology’, the module dealing with ‘educational practice’ was targeted. In this module, two texts were discussed. Through the first text written by Anderson (2005) the instructors introduce the ‘revised taxonomy of Bloom’ and aim to provide learners with insights about: (1) the importance of learning objectives, (2) the difficulties with regard to the formulation of such objectives, (3) the differences between the initial and the revised taxonomy of Bloom, (4) the link of assessment and instruction with the taxonomy, and (5) the revised taxonomy's potential application. The second text written by Mayer (2004) is used by the instructors to evoke learners' reflection on (1) the difference between ‘pure discovery learning’ and 'guided discovery learning', (2) research on both, and (3) the implications for education.

2.2.2. Module description

The module 'educational practice' was organized in the second semester of the 2017–2018 academic year. The module was provided in a blended learning format and consisted firstly of an online learning module in Moodle. Between two face-to-face contact sessions, learners had 28 days to progress through the environment and study the texts. In addition, learners were invited to participate in a 2 h-long contact sessions dealing with the content of the online learning module after having studied in Moodle.

2.3. The blended learning format

2.3.1. Experimental and control environments

The experimental and control environments were identical (except for the cues for calibration). With regard to the online component of the blended learning environment, a Moodle course was developed in a co-design fashion between instructors and researcher. For each text, the outline was identical and consisted of the following elements: (1) Objective of the text, (2) introduction (examples from practice), (3) text, (4) exercises, and (5) self-tests. With regard to the exercises and self-tests, each section of the texts was supported by practice exercises (not obligatory), preparing the learners for a self-test about the section addressed. Following each practice exercise an obligatory self-test was provided. After this test was submitted, learners could progress to the next section. Learners were allowed to choose a text to start with. During the studying of the texts learner control was limited, but after having finished the entire online learning module, learners could navigate freely through it. Also a (6) discussion forum was provided for course related discussions. In the experimental conditions cues for calibration were added to each section of the texts, to be more precisely before (in the form of a JOL-cue) and after (in the form of validity feedback) each self-test. With regard to the latter only functional validity feedback was given to learners in the functional validity feedback condition (F-condition), whereas learners in the functional and cognitive validity condition (FC-condition) received both functional and cognitive validity feedback. Fig. 1 shows the design of each condition.

For each of the elements in the online learning module learners' interactions were tracked. The structure of the face-to-face contact moment was identical for all. Learners were divided in groups of 5–6 and were instructed to make (1) a concept map of both texts, (2) elaborate a multimedia knowledge clip explaining the elements of the concept map and their link with practice, and finally they were asked to come up with (3) exam questions targeting the different levels of the 'revised taxonomy of Bloom' and focussing on the content of the two texts. The results of these exercises were discussed in the entire group and after the sessions, all materials were made available to all learners for further study.

2.3.2. Cues for calibration

2.3.2.1. Judgment of learning cues. Both experimental conditions contained judgement of learning cues to generate input for the automatization of the validity feedback. In line with Schraw (2009) these cues provided the learners with a question about their expected performance on the upcoming self-test, potentially based on the practise exercises. The cues were embedded and so implemented in line with guidelines on cue-use as formulated by Clarebout and Elen (2009). Each of the cues was embedded in the environment at the same level of the other content items and presented prior to each self-test. Learners in the experimental conditions were obliged to estimate their score. The operationalization of this design in the online learning module was done by using tests, including the 'test' icon in Moodle. Each page looked similar. In the face-to-face contact moment, no cues were provided. A typical judgement of learning cue (JOL-cue) asks learners for example: "How many of the four multiple choice questions dealing with the introduction of the Anderson (2005) text you think you will answer correctly?" or "You just studied the Mayer (2004) text, how many of the five multiple choice questions related to the content of the abstract you think you will answer correctly?". Based on this type of questions and learners' actual performance functional and cognitive validity feedback can be provided.

2.3.2.2. Functional and cognitive validity feedback cues. Both the functional and cognitive validity feedback cues were developed in line with Balzer and Doherty (1989) and Butler and Winne (1995). The functional validly feedback cues focused on the relationship between learners' judgement of learning (absolute accuracy index) obtained via the JOL-cues and their actual performance (obtained

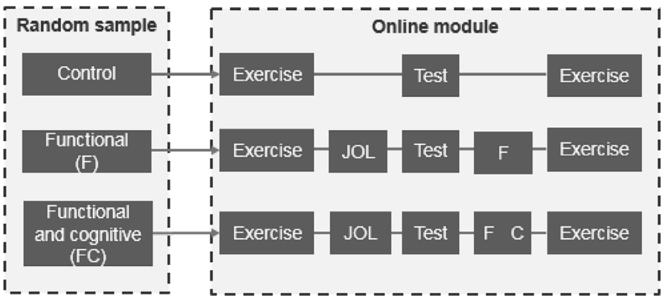

Fig. 1. Visual representation of the intervention for each condition.

via self-test scores). The accuracy of learners' judgement of learning was calculated using the absolute accuracy index formula proposed by Schraw (2009). This index was used as an indicator for assigning the appropriate functional validity feedback cue. This feedback contains any of the following labels: 'underestimate', 'estimate accurate', or 'overestimate' to the accuracy of prediction. Additionally learners were informed about their score on each self-test. A comparison of a learner's score was compared with the maximal score possible on the different self-tests. This was done by assigning either the label 'below 50% correct' or '50% or above 50% correct'. Combined this resulted in a personalized message for the learners stating: "You seem to [judgement of learning label] your ability and your score is [score label]". An example of a typical procedure for providing functional validity feedback cues is that learners first would have been presented a JOL-cue. Secondly learners would have been asked to complete four multiple choice questions about the introduction of the Anderson (2005) text. In case they answered for example "four out of four" on the JOL-cue asking them about how many of the four multiple choice questions they expect to answer correctly and their actual score was one out of four on the actual test, the kind of functional validity feedback they received would be: "You seem to 'overestimate' your ability and your score is 'below 50% correct'". If they answered "two out of four" and their actual score was "three out of four", the functional validity feedback they received would be: "You seem to 'underestimate' your ability and your score is 'above 50% correct'".

The cognitive validity feedback cues pertained to learners' perceptions about the relationship between the instructional components and performance. Learners received information on the link between (1) the instructional components, (2) the cognitive and metacognitive strategies needed to use these components, and (3) on the potential impact of both on their performance. Both the functional and cognitive validity feedback cues were embedded in the environment at the same level of the other content items and provided after the completion of each self-test. Additionally, through a popup learners were informed that the content under the link of the cue might significantly help their learning. The operationalization of this design in the online learning module was done by using feedback forms, including the 'megaphone' icon in Moodle. Each page looked similar. In the face-to-face contact moment, no cues were provided. With regard to an example of the procedure to provide a cognitive validity feedback cue, learners first received functional validity feedback like: "You seem to 'overestimate' your ability and your score is 'below 50% correct'". Based on this feedback learners would receive a cue aiming to direct them to taking appropriate action based on this feedback. Such cognitive validity feedback cues for example included following elements: "When reviewing the Introduction of Anderson (2005), did you notice each heading represents a specific part of the text?", "Without knowing the different headings it might be difficult to be successful in answering the tests", and "Summarizing might be an appropriate strategy to tackle this issue". Another example is: "An abstract contains a short summary of the text, it is often advised to use this summary as a guideline while reading the entire text".

2.4. Data structure of the log files

Each action made by learners within the online course was registered resulting in a time stamped event (TSE) database with as column headers the time stamp of the action, personal identifier of the user, and event name. A data-driven approach was chosen. Prior to the event analysis, no recoding or transformations took place. The three conditions (F-condition, FC-condition, and control condition) were identical and included six standard event names (see Table 1). Each of these event names refers to an attribute of the environment. Data reported on the attributes hence refer to specific (series of) events. In the F-condition two additional attributes were available, namely 'judgement of learning (JOL)' and 'functional validity feedback (F-feedback)'. In the FC-condition both were available too and a ninth one was available, 'cognitive validity feedback (C-feedback)'. As in the investigation of the relationship between cues for calibration and learners' behaviour the aim was to identify which sub-sequences occurred significantly more in which condition, sub-sequences including judgement of learning, functional validity, and cognitive validity feedback cues were excluded from the analysis as they only occur in the experimental conditions.

Finally, in the face-to-face contact moments no trace data were gathered as all learners received the same instruction and did not interact with the Moodle environment.

Table 1
Actions traced in the online learning environment.

Attribute	*Description*
Course	Landing page of the course. On this page, learners found an overview of the entire course, links to each of the texts addressed, and links to the discussion forum.
Objective	Pages elaborating on the learning objectives aimed for by the different texts.
Text	Downloadable version of both texts addressed during the online learning module.
Forum	Discussion forum including the viewing of the forum, posting of information, and all other interactions related to the forum.
Exercise	Practise exercise for the support of learning prior to the self-test.
Self-test	Formative test on each section of the different texts under investigation. Obligatory to be able to progress to the next part of the online learning module.
JOL-cue	Formative test to estimate performance on the self-test for each section addressed in the different texts. Obligatory to be to progress to the next part of the online learning module.
F-feedback	Feedback page containing a functional validity feedback cue providing information and an open question (free to answer or not) aiming to evoke learners' calibration.
C-feedback	Feedback page containing a cognitive validity feedback cue providing information and an open question (free to answer or not) aiming to evoke learners' calibration.

2.5. Instruments

2.5.1. Prior domain knowledge and domain knowledge

During the pre-test phase a performance based prior domain knowledge test was administered to investigate learners' prior domain knowledge. This prior domain knowledge test containing ten multiple-choice questions (including an "I don't know option") represented the content of the module. The test consisted of questions related to both texts (five per text). The test was scored on ten points. The same test was used as post-test to measure learners' domain knowledge.

2.5.2. Goal orientation

Learners' goal orientation was measured by using the merged version of two questionnaires of Elliot and Church (1997) and Elliot and McGregor (2001) for measuring learners' goal orientation as constructed by Lust (2012). Whereas the initial questionnaire of Elliot and Church (1997) measured solely three dimensions of goal orientation (mastery approach, performance avoidance, and performance approach), the revised questionnaire Elliot and McGregor (2001) incorporated the fourth dimension of mastery avoidance as well. These two questionnaires were merged into one that contained 21 items (Mastery goal approach (MGA) (6 items), Mastery avoidance approach (MAA) (4 items), Performance goal approach (PGA) (5 items), Performance avoidance approach (PAA) (6 items)). Answers for the items were given on a 5-point Likert type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

2.5.3. Academic self-concept

The Academic self-concept (ASC) scale comprised two 10-item subscales: learning confidence and learning effort. The learning confidence subscale assessed learners' feelings and perceptions about their academic competence (Liu, Wang, & Parkins, 2005). Example items are 'I am good in most of my course subjects' and 'most of my classmates are smarter than I am' (negatively worded). The learning effort subscale assessed learners' commitment to and involvement and interest in schoolwork (Liu et al., 2005).

2.5.4. Judgement of learning

Learners' judgement of learning was used in two ways. On the one hand as a dependent variable, on the other hand as the input for the adaptive feedback in the experimental conditions. For both purposes learners' judgement of learning was calculated in accordance with Schraw (2009). Learners' absolute accuracy was measured, indicating the precision of a single estimation of future score compared to performance on a single test (Maki et al., 2005). With regard to judgement of learning as an dependent variable, right before the pre-test and post-test for domain knowledge all learners were asked to estimate their score on both domain knowledge tests. Based on the comparison with their performance on the domain knowledge, a pre-test and post-test judgement of learning score was calculated. In view of generating automated functional validity feedback (for both experimental conditions - see: Cues for calibration), learners' estimation of their future score was gathered through the judgement of learning cues proceeding each self-test. For both purposes, learners were asked to score their estimation of future score on an ordinal scale with ten point interval that range from 0/10 (0%) to 10/10 (100%). Their actual score was measured through a single test. Scores were recalculated to an identical ten point interval also ranging from 0/10 (0%) to 10/10 (100%). Following this measurement, the absolute accuracy index was calculated. The formula of this index can be found below:

$$\text{Absolute Accuracy Index} = \frac{1}{N}\sum_{i=1}^{N}(c_i - p_i)^2$$

In this formula, c_i corresponds to the estimation of future score and p_i corresponds to the actual score. Each deviation score between learners' estimation of future score and actual score is squared so it ranges from zero to one, where a score of zero corresponds to perfect calibration accuracy and a score of one corresponds to no accuracy. Smaller deviations correspond to better accuracy.

2.5.5. Metacognitive skilfulness

In line with the literature on cues for calibration, we investigated learners' metacognitive skilfulness to get more profound insights in the cues' effect on learners' self-regulated learning. This was done by using an online aptitude measurement developed by Veenman (e.g., Veenman, Bavelaar, De Wolf, & Van Haaren, 2014; Veenman, Wilhelm, & Beishuizen, 2004), called "The otter task". This measurement is a computerized learning-by-discovery task in Authorware. The otter-task requires learners to experiment with five independent variables in order to discover their (combined) effects on the growth of the otter population. The five variables were habitat, environmental pollution, public entrance, setting out new otter couples, and feeding fish in wintertime. Independent variables could have no effect on the otter population (public entrance), a main effect (habitat; pollution), and interact with another variable (habitat x setting out otter couples; pollution x feeding fish). For each experiment, participants could choose a value for the five variables by clicking on the pictograms on the left, and then order the computer to calculate the growth of the otter population. Results of experiments done were transferred to a storehouse where learners could scroll up and down to consult earlier results. After a minimum of fifteen experiments, an exit button occurs which allows the learners to leave "The otter task", nonetheless they are free to continue. All actions done by the learners are logged in a text file. This log file is scored for metacognitive skilfulness through ten log file indicators, namely: (1) number of experiments, (2) think time, (3) scroll down, (4) scroll up, (5) transition with one altered variable, (6) mean number of changes, (7) number of unique experiments, (8) variation of variables, (9) systematic changes, and (10) complete variation of variables. All learners' values per log file indicator were standardized into z-scores. Finally, mean z-scores were

Table 2
Pre and post reliability analysis per construct.

Latent variable	*Construct*	*α pre*	*α post*
Cognition	Domain knowledge (DK) (10 items)	.65	.66
Goal orientation	Mastery goal approach (MGA) (6 items)	.79	.75
	Mastery avoidance approach (MAA) (4 items)	.73	.80
	Performance goal approach (PGA) (5 items)	.86	.86
	Performance avoidance approach (PAA) (6 items)	.74	.77
Academic self-concept	Learning effort (LE) (10 items)	.79	.80
	Learning confidence (LC) (7 items)	.74	.72
Metacognition	Metacognitive skilfulness (MS) (10 log file indicators)	.84	

calculated over the ten log file indicators as an overall measure of metacognitive skilfulness (for a full account of the methodology, see: Veenman et al. (2014)). This calculation resulted in an individual metacognitive skilfulness score per learner, comparing learners' individual score with the sample score.

2.5.6. The quality of the instruments

Traditional reliability analysis (Cronbach's alpha) was used in order to investigate the quality of the measurement instruments. Table 2 depicts the Cronbach's alpha values of the different scales. Given the threshold of 0.70 as proposed by Nunnally and Bernstein (1994), all instruments seem to be in reach of this threshold.

2.6. Procedure

Learners' were randomly assigned to three separate but identical learning environments, either the control group, the F-condition, or the FC-condition. All (n = 151) learners attending the module were invited to complete the otter task (during four available timeslots) prior to their first login in the online learning module. The learners got 60 min to complete this task. The online pre-test questionnaire, the pre-test judgement of learning question, and prior domain knowledge test were administered at the start of the online learning module and obligatory to activate the content of the online learning module. Learners got 28 days' time to complete the online learning module, learners in the experimental conditions received cues for calibration during that time; learners in the control condition did not. After the completion of the intervention, learners in the three conditions completed the online post-test questionnaire, the post-test judgement of learning question, and the domain knowledge test. The learners did not receive any other form of instruction on the module content during the time period between the pre-test and the post-test. For the matching of the different datasets anonymized student IDs were used.

2.7. Analysis

First we quartered the learners based on their metacognitive skilfulness score. This was by done by ordering all learners' scores from the lowest to the highest, followed by dividing them in 4 groups. Each group represented 25% of the sample. Learners were assigned a quartile number indication in which quartile their score was situated (1 = low to 4 = high). In this way, a new categorical variable metacognitive skilfulness quartile membership (PMSQ) was created and used as an independent variable throughout the analyses. Secondly, descriptive statistics were calculated for each of the pre-test variables and both pre-test and post-test variables were investigated for correlations, to identify the need for multivariate or univariate tests. Third and final, through a one-way multivariate analysis of variance (MANOVA) with condition as independent variable and the pre-test variables as dependent variables, followed by univariate analyses of variance (ANOVAs) the three conditions' comparability among each other was checked for learners' prior domain knowledge, goal-orientation, academic self-concept, and metacognitive skilfulness quartile membership.

2.7.1. Investigation of learning behaviour

The event sequence analysis consisted of two major steps (e.g., Cicchinelli et al., 2018; Zhou, 2016, pp. 1007–1012). First frequent event sub-sequences were identified using exploratory sequence analysis. Secondly, discriminant frequent event sub-sequences were identified by using an explanatory approach as well. The latter analysis was based on the condition learners were in. This to identify what sub-sequences (dependent variable) occurred significantly more in which condition (independent variable). A similar approach was adopted for metacognitive skilfulness (PMSQ) and for the interaction between condition and metacognitive skilfulness. The learners' behavioural data was imported in R-statistics and analysed using the TraMineR package (Gabadinho, Ritschard, Mueller, & Studer, 2011).

A similar approach to identify frequent event sub-sequences was used as Jovanović, Gašević, Dawson, Pardo, and Mirriahi (2017) and Van Laer and Elen (2016, pp. 1–31). Both studies emphasize the importance of two parameters when identifying frequent event sub-sequences. The first one is the time constraint (Studer, Mueller, Ritschard, & Gabadinho, 2010). As we followed a data-driven approach while investigating the ecological order of events, we chose to set this parameter on one. This indicates that only events that actually occurred following each other are included. Events further apart in time are not considered. The second one is the relative threshold number of times (pMinSupport) a sub-sequence occurs among the different learners (Müller, Studer, Gabadinho, &

Ritschard, 2010). In this study, this parameter was arbitrarily set on 0.25 to assure frequent sub-sequences occurred at least in 25% of the learners.

Discriminant frequent event sub-sequence were identified in line with Kim and Shute (2015) and with Grover et al. (2017). The significant discriminating ability of the sub-sequences was first based on differences between conditions learners were in, secondly on metacognitive skilfulness, and finally on the interaction of the condition learners were in and learners' metacognitive skilfulness quartile membership (PMSQ). To be able to calculate the discriminating abilities of a frequent sub-sequence two arguments are needed (a) a sub-sequence (subseq) object containing the sub-sequences considered for discriminating the groups and (b) the variable that defines the groups (groups) (Garza, 2016). A chi-square test is used to investigate the significance of the relationship between the observed and expected occurrence of a frequent sub-sequence for each value of the measured variables (Studer et al., 2010). Finally, the effect sizes are calculated using the Cramer's V. The Cramer's V expresses the relationship between a certain discriminating frequent sub-sequence and the learners' characteristics and is reported in a value between zero and one. The closer to one the higher the relation. Cohen (1988) refers to small (≤ 0.30), medium (≥ 0.30 and $\leq$ 50), and large (≥ 0.50) effect sizes.

2.7.2. *Investigation of learning outcomes*

In order to (1) examine the effect of the instructional intervention on learners' learning outcomes and (2) examine the interaction effect of instructional intervention and learners' metacognitive skilfulness on learners' learning outcomes, a two-way multivariate analysis of covariance (MANCOVA) test with pre-test and post-test data will be done with as dependent variables learners' post-test scores on: domain knowledge (DK), mastery goal approach (MGA), mastery avoidance approach (MAA), performance goal approach (PGA), performance avoidance approach (PAA), learning effort (LE) learning confidence (LC), and judgement of learning (JOL). Condition and learners' metacognitive skilfulness (PMSQ) will be used as independent variables. All learners' pre-test scores are used as covariates. The main effects, followed by the interaction effects, and univariate tests will be reported. Additionally, pairwise comparison with a Bonferroni correction will further investigate the found effects. Nonetheless, before conducting the MANCOVA test, the variables were tested for normality (Shapiro–Wilks' test), sphericity (Mauchly's Test of Sphericity), and homogeneity of variances (Levene's test). Fig. 2, visualizes the MANCOVA used.

3. Results

3.1. *Descriptive statistics*

151 learners were included in the analysis and were assigned to a quartile, based on their score on metacognitive skilfulness. The first quartile (n = 38) represented Z-scores between −1.26 and −0.61 ($M = -0.91$, $SD = 0.208$), the second quartile (n = 38) between −0.60 and −0.15 ($M = -0.39$, $SD = 0.157$), the third quartile (n = 39) between −0.14 and 0.36 ($M = 0.11$, $SD = 0.160$), and the fourth quartile (n = 37) between 0.39 and 5.38 ($M = 1.36$, $SD = 1.121$). Descriptive statistics of learners' pre-test scores of the dependent variables can be found in Table 3.

3.2. *Pre-test comparison of the experimental and control conditions*

All 151 learners (control = 49, F-condition = 48, and FC-condition = 54) participated in the otter task, the prior judgement of learning question, the prior domain-knowledge test, and the pre-test questionnaire. The pre-test scores correlated weakly to moderately, so a multivariate analysis of variance (MANOVA) was applied to compare learners' pre-test scores (as dependent variables) for the three condition (independent variable). The MANOVA showed no significant differences (F (16, 282) = 1.59, Wilk's $\Lambda = 0.933$, $p = .71$) for experimental and control conditions. Nonetheless, the univariate tests showed a significant differences for domain knowledge (DK) (F (2, 151) = 3.42, $p = .035$) among the different conditions. Learners in the F-condition seemed to score significantly ($p = .012$) higher ($MD = 1.15$) than learners' in the FC-condition (see Table 4). No difference was found for the other variables. Also with regard to learners' metacognitive skilfulness quartile membership (PMSQ) no difference in occurrence between conditions could be found (F (2, 151) = 0.61, $p = .544$). Table 4 elaborates on the ANOVA tests and provides the key figures for each of the variables per condition. Table 5 presents the number of learners for learners' metacognitive skilfulness quartile membership (PMSQ) per condition.

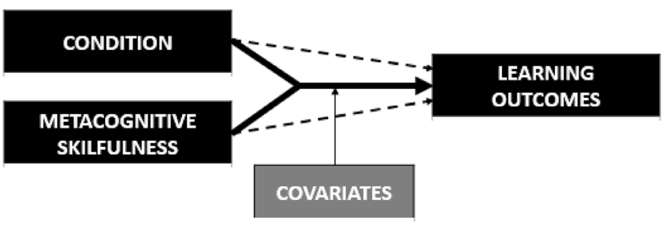

Fig. 2. Visual representation of the MANCOVA analysis.

Table 3
Descriptive statistics of the pre-test scores.

Pre-test variables	*N*	*Min*	*Max*	*M*	*S.E.*	*SD*	σ^2
Domain knowledge (PDK)	151	0.00	10.00	4.48	0.188	2.306	5.318
Mastery goal approach (MGA)	151	2.00	5.00	4.07	0.047	0.575	0.331
Mastery avoidance approach (MAA)	151	1.00	5.00	3.68	0.062	0.759	0.577
Performance goal approach (PGA)	151	1.00	4.20	2.37	0.060	0.751	0.564
Performance avoidance approach (PAA)	151	1.80	5.00	3.79	0.051	0.633	0.401
Learning confidence (LC)	151	1.57	4.43	3.02	0.040	0.498	0.249
Learning effort (LE)	151	1.90	4.90	3.45	0.041	0.510	0.261
Judgement of learning (JOL)	151	0.00	0.49	0.07	0.007	0.087	0.008

Correlation analysis showed weak (0.20-0.39) to moderate (0.40-0.59) correlations (Evans, 1996) between the different post-test variables, namely between PGA and LC ($r(149) = 0.36, p < .001$), PAA and MGA ($r(149) = 0.16, p = .043$), PAA and MAA ($r(149) = 0.51, p < .001$), PAA and LC ($r(149) = -0.16, p < .01$), and MAA and LE ($r(149) = 0.38, p < .001$). No other correlations were found.

Table 4
Comparison of the pre-test scores per condition.

Pre-test variables	Condition	*N*	*Min*	*Max*	*M*	*S.E.*	*SD*	σ^2
Domain knowledge (PDK)	Control	49	0.00	9.00	4.63	0.346	2.421	5.862
	F-condition*	48*	0.00*	10.00*	5.02*	0.354*	2.454*	6.021*
	FC-condition	54	0.00	8.00	3.87	0.263	1.933	3.738
	Total	151	0.00	10.00	4.48	0.188	2.306	5.318
Mastery goal approach (MGA)	Control	49	2.00	5.00	4.06	0.094	0.660	0.436
	F-condition	48	3.25	5.00	4.08	0.057	0.393	0.155
	FC-condition	54	2.50	5.00	4.10	0.087	0.641	0.411
	Total	151	2.00	5.00	4.08	0.047	0.577	0.333
Mastery avoidance approach (MAA)	Control	49	2.67	5.00	3.86	0.101	0.707	0.500
	F-condition	48	1.00	5.00	3.63	0.122	0.843	0.710
	FC-condition	54	1.67	5.00	3.57	0.098	0.717	0.514
	Total	151	1.00	5.00	3.68	0.062	0.761	0.580
Performance goal approach (PGA)	Control	49	1.20	4.00	2.44	0.090	0.632	0.399
	F-condition	48	1.00	4.20	2.23	0.119	0.825	0.681
	FC-condition	54	1.00	4.20	2.46	0.105	0.775	0.600
	Total	151	1.00	4.20	2.38	0.061	0.751	0.564
Performance avoidance approach (PAA)	Control	49	1.80	5.00	3.75	0.094	0.659	0.435
	F-condition	48	2.00	5.00	3.73	0.103	0.713	0.508
	FC-condition	54	2.60	5.00	3.90	0.072	0.527	0.277
	Total	151	1.80	5.00	3.80	0.052	0.635	0.403
Learning confidence (LC)	Control	49	1.86	3.86	2.98	0.060	0.418	0.175
	F-condition	48	2.00	4.14	3.00	0.082	0.568	0.323
	FC-condition	54	1.57	4.43	3.07	0.069	0.509	0.259
	Total	151	1.57	4.43	3.02	0.041	0.500	0.250
Learning effort (LE)	Control	49	2.50	4.60	3.46	0.070	0.490	0.240
	F-condition	48	2.60	4.30	3.50	0.061	0.421	0.177
	FC-condition	54	1.90	4.90	3.40	0.082	0.600	0.359
	Total	151	1.90	4.90	3.45	0.042	0.511	0.261
Judgement of learning (JOL)	Control	49	0.00	0.36	0.06	0.011	0.076	0.006
	F-condition	48	0.00	0.49	0.08	0.016	0.109	0.012
	FC-condition	54	0.00	0.36	0.06	0.010	0.075	0.006
	Total	151	0.00	0.49	0.07	0.007	0.087	0.008

*significant difference in score.

Table 5
The number for learners' metacognitive skilfulness quartile membership (PMSQ) per condition.

Condition	*PMSQ 1*	*PMSQ 2*	*PMSQ 3*	*PMSQ 4*	Total
Control	16	11	11	11	49
Experiment F	10	13	12	13	48
Experiment FC	12	14	15	13	54
Total	38	38	38	37	151

3.3. The effect of condition and metacognitive skilfulness on learners' learning behaviour

3.3.1. Condition

The learners (n = 149) included in the event sequence analysis generated 54434 events over the timespan of 28 days. A total of 249 frequent sub-sequences were extracted (TimeGap = 1; pMinSupport = .25).18 significant discriminant sub-sequences (pValueLimit = .05) were identified. Sub-sequences contained between two and seven events.

Learners in the control condition made significantly the most use of sub-sequences consisting of Self-test events followed by other Self-test events (between $\chi(2) = 83.848$, $p < .001$, $V = 0.56$ and $\chi(1) = 98.706$, $p < .001$, $V = 0.66$). Whereas for the control condition the standardized residuals scores were between 6.50 and 7.00, for the F-condition they were between −2.97 and −3.35, and for the FC-condition between −3.37 and −3.51. Examination of the Cramer's V scores indicate, according to Cohen (1988), large effects (≥0.50).

Learners in the control condition also made significantly the most use of sub-sequences consisting of Self-test events followed by Exercises events. Here chi-square tests (between $\chi(2) = 11.964$, $p < .001$, $V = 0.08$ and $\chi(1) = 16.092$, $p < .001$, $V = 0.11$) showed smaller effect sizes (≤0.30). The standardized residuals for the control condition were between 2.22 and 2.50, for the F-condition between −0.04 and −0.42, and for the FC-condition between −1.96 and −2.33.

Learners in the F-condition (*SR* between 0.51 and 1.72) used significantly more sub-sequences related to Exercise events followed by other Exercise events (χ^2-tests between $\chi(2) = 6.004$, $p < .001$, $V = 0.04$ and $\chi(1) = 6.546$, $p < .001$, $V = 0.05$). The control condition's standardized residuals score were between −1.23 and 0.82 and the FC-condition's between −1.35 and −0.15.

Finally, learners in the FC-condition did never demonstrated significantly more sub-sequences to any types of events for all significant discriminant sub-sequences.

3.3.2. Metacognitive skilfulness

In line with the investigation of the effect of condition on learners' learning behaviour, the effect of metacognitive skilfulness (PMSQ) was also studied. Significant discriminant sub-sequences, based on learners' metacognitive skilfulness (PMSQ) (quartile 1 to quartile 4) were identified. The same analysis was applied as when addressing the effect of the condition on learners' learning behaviour. Only 3 significant discriminant sub-sequences (pValueLimit = .05) were identified. Sub-sequence contained between four and six events.

The three discrimant sub-sequences, all showed Text events (downloading of one of the two articles learners needed to read) followed by Forum events. Learners belonging to the lowest quartile (Q1) used these sub-sequences significantly more events (between $\chi(3) = 7.814$, $p = .050$, $V = 0.23$ and $\chi(3) = 8.069$, $p = .044$, $V = 0.82$) than learners belonging to quartile 2 (*SR* between 0.22 and 0.36), quartile 3 (*SR* between −0.02 and −0.07), or quartile 4 (*SR* between −1.89 and −1.99). Examination of the Cramer's V scores indicate according to Cohen (1988) small effects of learners' metacognitive skilfulness on learners' learning behaviour (≥0.50).

3.3.3. Condition and metacognitive skilfulness

Finally, to investigate on the interaction of condition and metacognitive skilfulness (PMSQ) on learners' learning behaviour, significant discriminant sub-sequences based on condition and metacognitive skilfulness (PMSQ) (1 = low to 4 = high) were identified. 10 significant discriminant sub-sequences (pValueLimit = .05) were identified and compared among 12 groups (3 conditions x 4 quartiles). Sub-sequence contained between three and six events.

Learners in the control condition belonging to the lowest quartile (Q1) of metacognitive skilfulness made significantly the most use of sub-sequences consisting of Self-test events followed by other Self-test events (between $\chi(11) = 86.47$, $p < .001$, $V = 0.76$ and $\chi(11) = 101.280$, $p < .001$, $V = 0.82$). Examination of the Cramer's V scores indicate according to Cohen (1988) large effects based on condition and PMSQ (≥0.50). Here standardized residuals score were between 4.41 and 4.68. Their counterpart belonging to different quartiles of PMSQ, but to the same condition, used fewer such sub-sequences. For learners belonging to the second quartile (Q2), standardized residuals score were between 2.50 and 3.75, for learners belonging to the third quartile (Q3), between 2.94 and 3.41, and for the highest quartile (Q4), between 2.36 and 2.64. The same observation could be made for learners belonging to the lowest quartile (Q1) in the F-condition (*SR* between −0.97 and −1.64) and the FC-condition (*SR* between −1.75 and −1.88). Learners belonging to higher quartiles in the F-condition or the FC-condition made less use of sub-sequences consisting of Self-test events followed by other Self-test events. Learners in the FC-condition belonging to the second quartile (Q2) used the least of these sub-sequences (SR between −1.89 and −2.03).

Learners in the F condition belonging to highest quartile (Q4) made significantly more use of sub-sequences consisting of Self-test events followed by Exercises events. Here Chi squared tests (between $\chi(11) = 20.774$, $p = .036$, $V = 0.37$ and $\chi(11) = 23.431$, $p = .015$, $V = 0.40$) showed medium effect sizes (between 0.30 and 0.50). The standardized residuals were between 0.83 and 1.16, where they were for their counterparts belonging to the third quartile (Q3), between −0.83 and −1.27, for learners belonging to the second quartile (Q2), between 0.38 and 0.75, and for the lowest quartile (Q1), between −0.97 and −1.26. The same observation with regard to sub-sequences consisting of Self-test events followed by Exercises events could be made for learners belonging to highest quartile (Q4) of metacognitive skilfulness in the control condition (*SR* between 1.25 and 1.71), but not for learners belonging to this quartile in the FC-condition (*SR* between −1.75 and −1.99). Learners belonging to lower quartiles in the control condition or the FC-condition made less use of sub-sequences consisting of Self-test events followed by Exercise events. Learners in the FC-condition belonging to the highest quartile (Q4) used the least of these sub-sequences (SR between −1.75 and −1.99).

Finally, learners in the highest quartile (Q4) and the FC-condition did never demonstrated significantly more sub-sequences to any

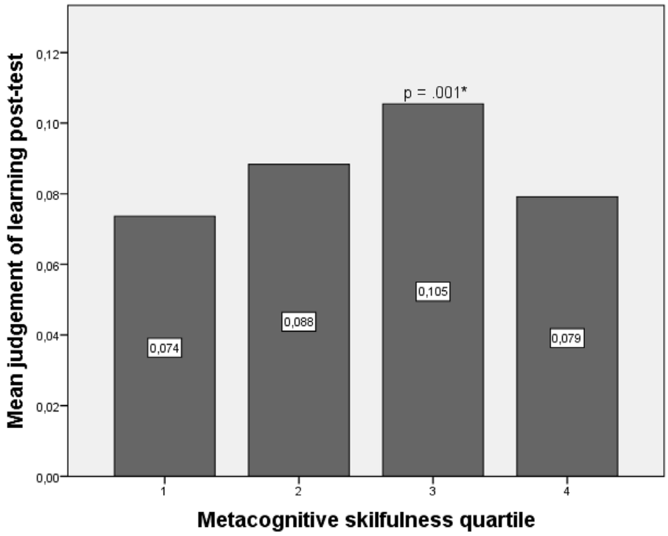

Fig. 3. Mean judgement of learning post-test scores per metacognitive skilfulness quartile.

types of events for all significant discriminant sub-sequences.

3.4. The effect of condition and metacognitive skilfulness on learners' learning outcomes

For the multivariate tests, the main effect of condition (F (16, 150) = 0.908, p = .562, Wilk's Λ = 0.831, η_p^2 = 0.09) was not significant, indicating that condition had no direct effect on the dependent variables under investigation. In contrast to this, the main effect of learners' pre-test metacognitive skilfulness (PMSQ) was significant (F (24, 218.124) = 1.987, p = .005, Wilk's Λ = 0.564, η_p^2 = 0.17), indicating that a different degree of skilfulness affects the dependent variables. Here, univariate tests showed that only learners' judgement of learning (F (3, 138) = 6.025, p = .001, η_p^2 = 0.18) significantly differed depending on learners' degree of skilfulness. Pairwise comparisons using a Bonferroni correction showed that learners in the third quartile (Q3) (M = 0.105) scored significantly less accurate, than learners in the lowest quartile (Q1) (MD = −0.03, SE = 0.025, p = .023), the second quartile (Q2) (MD = −0.02, SE = 0.025, p = .022), or in the highest quartile (Q4) (MD = −0.03, SE = 0.025, p = .011). Fig. 3 shows the mean post-test scores for judgement of learning per metacognitive skilfulness quartile.

Results also reveal a significant interaction effect between condition and metacognitive skilfulness (F (48, 373.094) = 1.560, p = .025, Wilk's Λ = 0.591, η_p^2 = 0.09). The univariate tests showed learners' judgement of learning (F (6, 138) = 3.862, p = .002, η_p^2 = 0.13) and learning confidence (F (6, 138) = 1.474, p = .017, η_p^2 = 0.17) significantly differed depending on the condition learners were in and learners' degree of metacognitive skilfulness.

Post-hoc comparison using a Bonferroni correction showed that learners belonging to quartile 3 in the FC-condition (M = 0.25) scored significant less accurate than learners in the control condition (MD = −0.11, SE = 0.051, p = .103) or F-condition (MD = −0.17, SE = 0.042, p < .001) belonging to the same quartile, indicating that they were less accurate. The opposite was found for learners belonging to the highest quartile (Q4) in the FC-condition (M = 0.04). Learners belonging to the control condition (MD = 0.10, SE = 0.048, p = .033) or the F-condition (MD = 0.01, SE = 0.042, p = .753) and the same quartile, scored less accurate. The other conditions and quartiles configurations did not show any significant differences with regard to learners' judgement of learning. Fig. 4 shows the estimate marginal means for judgement of learning per condition and metacognitive skilfulness quartile.

Finally, for learning confidence, post-hoc comparisons using a Bonferroni correction showed that learners with the lowest degree of metacognitive skilfulness (Q1) in the F-condition (M = 2.98) scored lowest on learning confidence. Learners belonging to the same quartile in the control condition (MD = 0.25, SE = 0.125, p = .050) or FC-condition (MD = 0.03, SE = 0.108, p = .804) scored higher on learning confidence. The opposite was found for learners belonging to the highest quartile (Q4) in the F-condition (M = 3.40) who scored highest on learning confidence. Learners belonging to the same quartile in the control condition (MD = −0.37, SE = 0.137, p = .009) or FC-condition (MD = −0.18, SE = 0.129, p = .178) scored lower on learning confidence. No other significances were found in the univariate tests for learning confidence. Fig. 5 shows the estimate marginal means for learning confidence per condition and metacognitive skilfulness quartile.

4. Discussion

4.1. Findings

4.1.1. Hypothesis 1: "Cues for calibration affect learners' learning behaviour."

Cues for calibration clearly seem to affect learners' learning behaviour, so hypothesis one could be confirmed. Learners in the

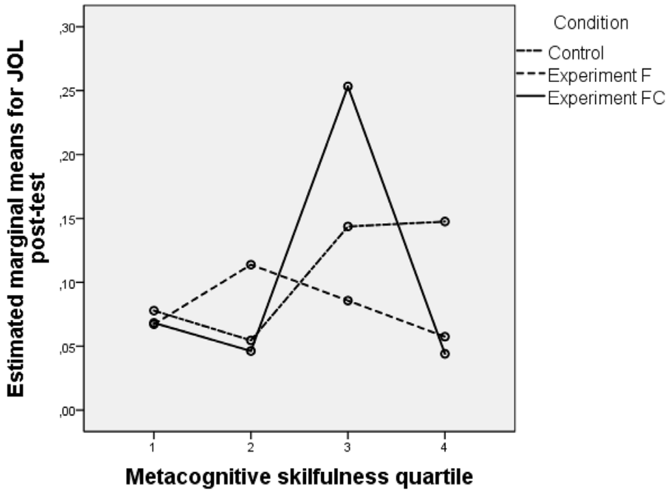

Fig. 4. Estimate marginal means for judgement of learning per condition and metacognitive skilfulness quartile.

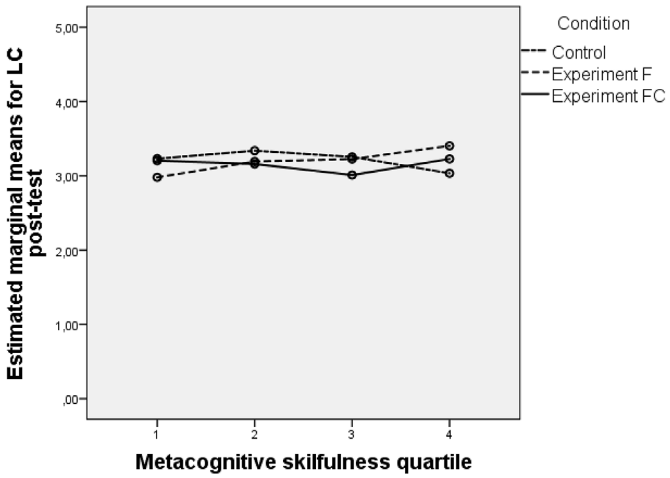

Fig. 5. Mean judgement of learning post-test scores per condition and metacognitive skilfulness quartile.

control condition made significantly more use of sub-sequences consisting of Self-test events followed by other Self-test events compared to the learners in the experimental conditions. Learners in the F-condition seemed to use such sub-sequences significantly less, and even less so by learners in the FC-condition. Similar findings were found in relation to sub-sequences consisting of Exercises events following Self-test events. Learners in the F-condition only used significantly more sub-sequences related to Exercise events following other Exercise events compared to the learners in the control condition, and learners in the FC-condition who used the sub-sequences the least. For all 18 significant discriminant sub-sequences learners in the FC-condition, never demonstrated significantly more use of sub-sequences compared to the other two conditions. Additionally, independent from the condition also the effect of learners' metacognitive skilfulness was investigated. Here there seemed to be hardly any differences among learners. Learners with the lowest degree of metacognitive skilfulness (Q1) made significantly more use of sub-sequences related to Text events followed by Forum events. The higher learners' score was for metacognitive skilfulness, the fewer they used these sub-sequences.

4.1.2. Hypothesis 2: "Cues for calibration affect learners' learning behaviour differently for learners with different levels of metacognitive skilfulness."

Based on the following elements Hypothesis 2 can be regarded to be confirmed. Learners in the control condition belonging to the lowest degree quartile (Q1) made the most use of sub-sequences consisting of Self-test events followed by other Self-test events compared to learners in the other groups. Learners in the same condition but with other degrees of skilfulness used these sub-sequences less frequently. Learners belonging to the same quartile in the F-condition and the FC-condition exhibited the same behaviour. Learners belonging to different quartiles made less use of sub-sequences consisting of Self-test events followed by other Self-test events. Learners in the FC-condition belonging to the second quartile (Q2) used these sub-sequences the least.

Learners in the F condition belonging to the highest quartile (Q4) of metacognitive skilfulness made significantly more use of sub-

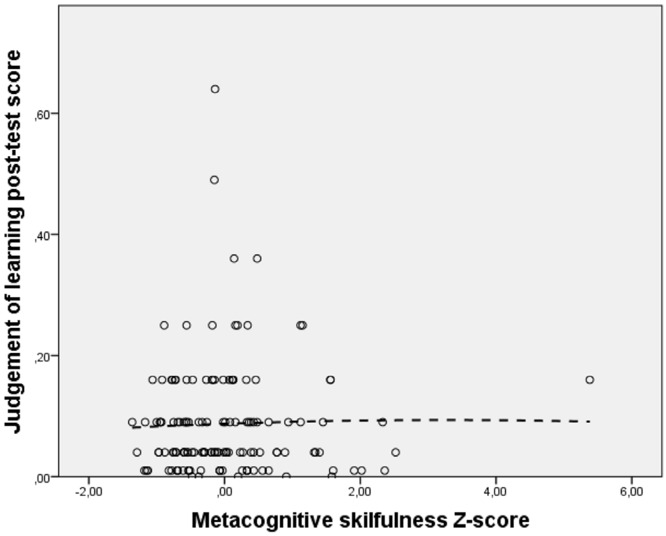

Fig. 6. Curvilinear relationship between metacognitive skilfulness and judgement of learning.

sequences consisting of Self-test events followed by Exercises events. Learners belonging to the highest quartile (Q4) in the control condition exhibited the same behaviour. Learners belonging to different quartiles in the control condition or the FC-condition made less use of sub-sequences consisting of Self-test events followed by Exercise events. Learners in the FC-condition belonging to quartile 4 used the least of these sub-sequences.

Finally, learners with the highest degree (Q4) of metacognitive skilfulness in the FC-condition did never demonstrated significantly more sub-sequences to any types of events for all significant discriminant sub-sequences.

4.1.3. Hypothesis 3: "Cues for calibration positively affect learners' learning outcomes."

No significant main effect of condition on learners' learning outcomes could be found. Therefore, Hypothesis 3 is falsified. However, learners' metacognitive skilfulness significantly affected learners' learning outcomes. Further univariate analyses have revealed this only to be the case for learners' judgement of learning. Learners with metacognitive skilfulness degrees between −0.14 and 0.36 (Q3) judged their learning significantly less accurate on the post-test than learners with other levels of metacognitive skilfulness. Explorative analysis showed indications of a curvilinear relationship between metacognitive skilfulness and judgement of learning. Fig. 6 shows this relationship.

4.1.4. Hypothesis 4: "Cues for calibration positively affect learners' learning outcomes most when learners have high levels of metacognitive skilfulness."

Finally, based on the following elements, Hypothesis 4 can said to be confirmed although in an unexpected direction. Investigating the interaction effect of the cues for calibration (conditions) and learners' metacognitive skilfulness (PMSQ) on learners' learning behaviour, it became clear that certain constellations of condition and degrees of learners' metacognitive skilfulness significantly affected learners' learning outcomes. Nonetheless, this was the case only for two dependent variables. Univariate tests showed learners' judgement of learning and learning confidence to be affected by the interaction of both independent variables. With regard to judgement of learning, results showed that learners with metacognitive skilfulness degrees between −0.14 and 0.36 (Q3) in the FC-condition scored less accurate, than learners in the control condition or F-condition belonging to the same quartile. For learners with high metacognitive skilfulness degrees between 0.39 and 5.38 (Q4) of the FC-condition, the results showed the opposite. These learners were more accurate than the others were.

In relation to learning confidence, results showed that learners with low degrees of metacognitive skilfulness between −1.26 and −0.61 (Q1) in the F-condition scored lowest on learning confidences. Learners belonging to the same quartile in the control condition or FC-condition scored higher. The opposite result was found for learners with high degrees of metacognitive skilfulness between 0.39 and 5.38 (Q4) of the F-condition. These learners scored higher on learning confidence than learners in the same quartile for the other conditions.

4.2. Exploration of the unexpected findings

The study presented shows that cues for calibration, affected learners' learning behaviour and outcomes, and so self-regulated learning. However, the directions are unexpected. Below we provide a possible explanation.

4.2.1. Learners' cue use

In line with current research on the link between instructional interventions and learners' learning behaviour through log-file data (e.g., Rienties, Toetenel, & Bryan, 2015; Wolff, Zdrahal, Nikolov, & Pantucek, 2013), the type of calibration cue learners received influenced their learning behaviour. The observation that learners in the control condition made significantly more use of certain subsequences compared to the other conditions was rather striking as it contrasts with research reporting greater learner involvement with the learning environment and cues provided, when cues for calibration are provided (Szabo, Falkner, Knutas, & Dorodchi, 2018; Timmers, Walraven, & Veldkamp, 2015). However, from a self-regulated learning theory perspective a decrease in particular learning behaviour might be explained as follows. When learners are capable to identify the instructional requirements set, to comply with them, and so to be successful in achieving the learning outcomes targeted, they are self-regulated learners (e.g., Wolters, Won, & Hussain, 2017). In line with this reasoning the fewer actions needed to achieve the outcomes targeted the more effective one's self-regulated learning is (Winne & Hadwin, 1998). When cues for calibration are provided, including functional and cognitive validity feedback (FC-condition), self-regulated learners direct their behaviour towards the information that helps them to achieve the learning outcomes targeted (e.g., Butler & Winne, 1995; Dunlosky & Thiede, 2013; Geitz, Joosten-ten Brinke, & Kirschner, 2016; Rienties & Toetenel, 2016). When instead learners are only provided with indications about their calibration efforts (F-condition) they might act on this feedback and adapt their behaviour by for example making more exercises in an attempt to progress (e.g., Tempelaar, Rienties, & Giesbers, 2015). Finally, when in contrast learners are not provided with any information about their calibration efforts (control condition), it is solely up to them to gather this information, potentially resulting in feedback-seeking behaviour (e.g., Harrison, Könings, Schuwirth, Wass, & van der Vleuten, 2015). In conclusion, as a result of providing learners with functional and cognitive validity feedback, highly metacognitive skilful learners might be selective and only engage in specific goal-directed behaviour, whereas learners struggling with controlling their learning might rather perform a plenty-fold of undirected behaviours (e.g., Fonseca, Martí, Redondo, Navarro, & Sánchez, 2014; Van Laer & Elen, 2016, pp. 1–31).

4.2.2. Learners' cue interpretation

As cues of calibration are inevitably interpreted through the lens of one's self-perceptions, it is important to understand how learners interpret the information provided to them (Eva et al., 2012). One way of doing this, is through the observation of changes in learners' learning outcomes. The findings of the study presented show that learners receiving functional validity feedback and having low degrees of metacognitive skilfulness, scored significantly lower for learning confidence in contrast to their counterparts in the control condition. According to research on the effect of cues for calibration on learners' learning confidence (e.g., Van der Kleij, Feskens, & Eggen, 2015), when learners are confronted with functional validity feedback, learners' might interpret the feedback for example as an indicator of personal failing or looming problems, rather than as a cue for them to re-calibrate (Hattie & Timperley, 2007). Especially when learners have low degrees of learning confidence this might be decisive for their further use of cues, as they might relate cues with negative experiences (Levine & Donitsa-Schmidt, 1998). In line with this reasoning, research points out that learners provided with functional and cognitive validity feedback do not have this problem, as cognitive validity feedback directs them to appropriate action to overcome this feeling (e.g., Ridder, McGaghie, Stokking, & Cate, 2015). Functional validity feedback did not only affect learners with low degrees of metacognitive skilfulness negatively, functional validity feedback also led to increased learning confidence for learners with high degrees of metacognitive skilfulness. Nonetheless, in the light of calibration, learning confidence without any performance related increase might lead to overestimation of one's own capabilities and further down the road to a decrease in performance (e.g., Dunlosky & Rawson, 2012).

4.2.3. Cue's effectiveness for increased performance

In line with our findings, Hellrung and Hartig (2013) present, in their systematic literature review, a substantial body of literature reporting increased learners' judgement of learning evoked by the use of functional and cognitive validity feedback. However, this was only the case for learners with high degrees of metacognitive skilfulness. One possible explanation for this is that learners with high degrees of metacognitive skilfulness are more aware of the different underlying strategies potentially supporting calibration and re-calibration (e.g., Hacker, Bol, & Bahbahani, 2008). This would result in more accurate estimations of one's performance (Callender, Franco-Watkins, & Roberts, 2016). So the information learners receive on the accuracy of their perceived level of performance in relation to their actual level, helps them to re-calibrate (e.g., Muis et al., 2016; Winne & Jamieson-Noel, 2002). Although for a change in accuracy to occur, learners need insight into the cognitive processes needed to calibrate their learning (e.g., Alexander, 2013; Dunlosky & Thiede, 2013). The combination of functional and cognitive validity feedback proved to provide calibration and showed increased judgement of learning. Based on the results of accurate calibration, learners monitor their learning and select cognitive and metacognitive strategies (e.g., error correction strategies, revision activities, etc.) which may help to proceed them in the direction of the desired level of performance (Narciss, 2017, pp. 173–189). The finding that certain learners increased in judgement of learning but not in performance might relate to the latter. Even when learners can calibrate external and internal feedback, they might not (1) possess or (2) be able to recall the cognitive and/or metacognitive strategies needed to act in a way that will produce increased performance in this specific occasion (e.g., Pintrich, 2002; Veenman, 1993). This would evoke sub-optimal self-regulated learning and hamper increased performance. Although we investigated the effect of learners' metacognitive skilfulness based on learners' domain general ability to control and apply cognitive and metacognitive processes, the cues for calibration provided might have lacked the potential to evoke the transfer of these processes to a domain specific context (Butler & Winne, 1995). This hypothesis has been supported by prior findings (e.g., Ardasheva et al., 2017; Dinsmore & Fryer, 2018) indicating that cues on the use of cognitive and/or metacognitive strategies should strongly align with the content provided (e.g., Alexander, 2018; Tricot & Sweller, 2014).

5. Further directions and conclusions

The present study documents fine-grained insights into the relationship between learners' self-regulated learning and cues for calibration. To obtain these insights, we first investigated learners' calibration-cue use based on learners' individual differences. Secondly, we operationalized self-regulated learning through learners' learning behaviour and outcomes. Investigating both learning behaviour and outcomes provides insights on learners' self-regulated learning, as well as on the nature of cues' effects. The current study reveals that differences in learner behaviour were related to condition and learners' metacognitive skilfulness, thus establishing a link between learners' self-regulated learning and cues for calibration. Finally, in the discussion of the results, we unravelled the effect of the design and content of the cues for calibration provided and hypothesized that when cues for calibration are provided through functional and cognitive validity feedback, learners' calibration capabilities will increase. Yet for this to result in goal-directed self-regulated learning and so increased achievement, we hypothesize learners might not only need to be supported in identifying and recalling the cognitive and metacognitive strategies needed, but also directed to how to apply the cognitive and metacognitive strategies in their context.

5.1. Further directions

Tapping in into the hypothesis that learners might not only need to be supported in identifying and recalling the cognitive and metacognitive strategies needed, but also need to be directed to how to apply these strategies in their specific context, a first direction for further research relates to the investigation of this hypothesis. By deploying a similar study design as the one presented in this manuscript, including on the one hand context general and on the other hand context specific cognitive validity feedback might allow this hypothesis might get tested. To further enrich our understanding, also some methodological challenges need to be addressed. A first challenge is the sample-size. A total of 151 learners were involved in the study – 48 in the control condition, 49 in the F-condition, and 54 in the FC-condition, for testing the interaction between condition and metacognitive skilfulness, more learners per condition might be advisable and so the power of some of the presented statistics might be debatable. A second challenge relates to the use of a data-driven approach to analyse learners' learning behaviour and its arbitrary parameter setting. In further research, this data-driven approach might be explored by experimenting with different parameter settings or a combination of data-driven and theory-driven approaches might be taken. With regard to the latter, this could be achieved for example by recoding events or sub-sequences based on a theoretical framework unrelated to self-regulation theory (for example a tool-use scheme). This would make the sub-sequences identified more meaningful and so interpretable. The third and final challenge relates to the relation between learners' self-regulated learning and the effect of cues. In this study the design of the cues and learners' metacognitive skilfulness were related to learners' learning behaviour. To be able to identify meaningful learning behaviour in the light of learning outcomes, future research might want to model the path of how different types of learners use the cues for calibration provide, leading them to certain learning outcomes for example through Hidden Markov Modelling.

5.2. Conclusions

Given the current lack of certainty regarding the effects of cues for calibration on learners' self-regulated learning, teachers and instructional designers remain dependent on inconsistent conceptual claims that cues for calibration may improve self-regulated learning. Studies such as the one presented here help both researchers and practitioners to distinguish between the effect of cues for calibration and how learners react to them. Establishing more fine-grained links between learners' characteristics, learning behaviour, and learning outcomes could help us propel the investigation of the effect of cues in intervention research.

Acknowledgements

We would like to acknowledge the support of the project "Adult Learners Online", financed by the Agency for Science and Technology (Project Number: SBO 140029), which made this research possible.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction, 24*, 1–3.
Alexander, P. A. (2018). Looking down the road: Future directions for research on depth and regulation of strategic processing. *British Journal of Educational Psychology, 88*(1), 152–166.
Anderson, L. W. (2005). Objectives, evaluation, and the improvement of education. *Studies In Educational Evaluation, 31*(2–3), 102–113.
Archer, J. C. (2010). State of the science in health professional education: Effective feedback. *Medical Education, 44*(1), 101–108.
Ardasheva, Y., Wang, Z., Adesope, O. O., & Valentine, J. C. (2017). Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes. *Review of Educational Research, 87*(3), 544–582. https://doi.org/10.3102/0034654316689135.
Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied, 24*(1), 43.
Arrastia-Chisholm, M. C., Torres, K. M., & Tackett, S. (2017). Using reflection to increase self-regulation among pre-service teachers. *Fostering Reflective Teaching Practice in Pre-Service Education, 148*.
Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds. *Instructional Science, 33*(5), 367–379.
Azevedo, R., Martin, S. A., Taub, M., Mudrick, N. V., Millar, G. C., & Grafsgaard, J. F. (2016). *Are pedagogical agents' external regulation effective in fostering learning with intelligent tutoring systems? Paper presented at the international conference on intelligent tutoring systems*.
Balzer, W. K., & Doherty, M. E. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106*(3), 410.

Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., & Gašević, D. (2017). Relevance of learning analytics to measure and support students' learning in adaptive educational technologies. *Paper presented at the proceedings of the seventh international learning analytics & knowledge conference*.
Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short-and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306.
Besser, E. D. (2016). *Exploring the role of feedback and its impact within a digital badge system from multiple perspectives: A case study of preservice teachers.* Purdue University.
Boelens, R., Van Laer, S., De Wever, B., & Elen, J. (2015). *Blended learning in adult education: Towards a definition of blended learning.* Brussels: Genth University.
Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69*(2), 133–151.
Broadbent, J., & Poon, W. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *Internet and Higher Education, 27*, 1–13.
de Bruin, A. B., & van Merriënboer, J. J. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction, 51*, 1–9.
Bui, T., & Loebbecke, C. (1996). Supporting cognitive feedback using system dynamics: A demand model of the global system of mobile telecommunication. *Decision Support Systems, 17*(2), 83–98.
Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.2307/1170684.
Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and learning, 11*(2), 215–235.
Chyung, S. Y. (1996). *The effects of conceptual tempo and computer-assisted instructional environment on college students' self-regulated learning and academic achievement.* Texas Tech University.
Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., et al. (2018). *Finding traces of self-regulated learning in activity streams*.
Clarebout, G., & Elen, J. (2009). The complexity of tool use in computer-based learning environments. *Instructional Science, 37*(5), 475. https://doi.org/10.1007/s11251-008-9068-3.
Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing self-regulation as a cyclical, context-specific phenomenon: Overview and analysis of SRL microanalytic protocols. *Educational Research International, 2012*. https://doi.org/10.1155/2012/428639.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, *Vol. 2*Hilsdale. *NJ*: Lawrence Earlbaum Associates.
Daniela, P. (2015). The relationship between self-regulation, motivation and performance at secondary school students. *Procedia-Social and Behavioral Sciences, 191*, 2549–2553. https://doi.org/10.1016/j.sbspro.2015.04.410.
Delgadillo, J., Overend, K., Lucock, M., Groom, M., Kirby, N., McMillan, D., et al. (2017). Improving the efficiency of psychological treatment using outcome feedback technology. *Behaviour Research and Therapy, 99*, 89–97.
DiFrancesca, D., Nietfeld, J. L., & Cao, L. (2016). A comparison of high and low achieving students on self-regulated learning variables. *Learning and Individual Differences, 45*, 228–236.
Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391–409.
Dinsmore, D. L., & Fryer, L. K. (2018). The intersection between depth and the regulation of strategy use. *British Journal of Educational Psychology, 88*(1), 1–8.
Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280.
Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24*, 58–61.
Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33*(1), 87–105.
Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153.
Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*(1), 218. https://doi.org/10.1037//0022-3514.72.1.218.
Elliot, A. J., & Dweck, C. S. (2013). *Handbook of competence and motivation.* Guilford Publications.
Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501.
Ellis, S. (2012). *Learning from errors: The role of after-event reviews Human fallibility.* Springer.
Endedijk, M. D., Brekelmans, M., Verloop, N., Sleegers, P. J., & Vermunt, J. D. (2014). Individual differences in student teachers' self-regulated learning: An examination of regulation configurations in relation to conceptions of learning to teach. *Learning and Individual Differences, 30*, 155–162. https://doi.org/10.1016/j.lindif.2013.12.005.
Ernst, B., & Steinhauser, M. (2017). Top-down control over feedback processing: The probability of valid feedback affects feedback-related brain activity. *Brain and Cognition, 115*, 33–40.
Eva, K. W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., et al. (2012). Factors influencing responsiveness to feedback: On the interplay between fear, confidence, and reasoning processes. *Advances in Health Sciences Education, 17*(1), 15–26.
Evans, P. (1996). Government action, social capital and development: Reviewing the evidence on synergy. *World Development, 24*(6), 1119–1132.
Fonseca, D., Martí, N., Redondo, E., Navarro, I., & Sánchez, A. (2014). Relationship between student profile, tool use, participation, and academic performance with the use of Augmented Reality technology for visualized architecture models. *Computers in Human Behavior, 31*, 434–445.
Frysak, J. (2017). Feedback mechanisms for decision support systems: A literature review. *Paper presented at the world conference on information systems and technologies*.
Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1–37.
Garza, M. M. M. (2016). *Coevolution of theory and data analytics of digital game-based learning.* Vanderbilt University.
Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71.
Geitz, G., Joosten-ten Brinke, D., & Kirschner, P. A. (2016). Changing learning behaviour: Self-efficacy and goal orientation in PBL groups in higher education. *International Journal of Educational Research, 75*, 146–158.
Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77*(3), 334–372. https://doi.org/10.3102/003465430303953.
Grover, S., Basu, S., Bienkowski, M., Eagle, M., Diana, N., & Stamper, J. (2017). A framework for using hypothesis-driven approaches to support data-driven learning analytics in measuring computational thinking in block-based programming environments. *ACM Transactions on Computing Education, 17*(3), 14. https://doi.org/10.1145/3105910.
Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and learning, 3*(2), 101–121.
Harrison, C. J., Könings, K. D., Schuwirth, L., Wass, V., & van der Vleuten, C. (2015). Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education, 20*(1), 229–245.
Hartley, A. A., Kieley, J. M., & Slabach, E. H. (1990). Age differences and similarities in the effects of cues and prompts. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 523.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
Hellrung, K., & Hartig, J. (2013). Understanding and using feedback–A review of empirical studies concerning feedback from external evaluations to teachers.

Educational Research Review, 9, 174–190.
Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education, 33*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001.
Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education, 87*, 340–356. https://doi.org/10.1016/j.compedu.2015.07.009.
Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly, 25*(2), 88–102.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.
Levine, T., & Donitsa-Schmidt, S. (1998). Computer use, confidence, attitudes, and knowledge: A causal analysis. *Computers in Human Behavior, 14*(1), 125–146.
Lin, B., Coburn, S. S., & Eisenberg, N. (2016). Self-regulation and reading achievement. *The cognitive development of reading and reading comprehension,* 67–86.
Liu, Z., Dev, H., Dontcheva, M., & Hoffman, M. (2016). *Mining, pruning and visualizing frequent patterns for temporal event sequence analysis*.
Liu, W. C., & Wang, C. K. J. (2005). Academic self-concept: A cross-sectional study of grade and gender differences in a Singapore secondary school. *Asia Pacific Education Review, 6*(1), 20–27. https://doi.org/10.1007/bf03024964.
Liu, W., Wang, C., & Parkins, E. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology, 75*(4), 567–586. https://doi.org/10.1348/000709905X42239.
Lust, G. (2012). *Opening the black box. Students' tool-use within a technology-enhanced learning environment: An ecological-valid approach*.
Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57*(3), 2135–2144. https://doi.org/10.1016/j.compedu.2011.05.010.
Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*(4), 723.
Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14.
Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2012). *Organizing for behavior change by collecting useful data. Behavior analysis for lasting change*.
Mory, E. H. (2004). Feedback research revisited. *Handbook of research on educational communications and technology, 2*, 745–783.
Muis, K. R., Winne, P. H., & Ranellucci, J. (2016). The role of calibration bias and performance feedback in achievement goal regulation. *International Education Research, 4*, 14–36.
Müller, N. S., Studer, M., Gabadinho, A., & Ritschard, G. (2010). Analyse de séquences d'événements avec TraMineR. *Paper presented at the EGC*.
Narciss, S. (2017). *Conditions and effects of feedback viewed through the lens of the Interactive Tutoring Feedback Model Scaling up assessment for learning in higher education.* Springer.
Nelson, T., Narens, L., & Bower, G. (1990). *The psychology of learning and motivation. Metamemory: A theoretical framework and new findings*.
Nunnally, J. C., & Bernstein, I. H. (1967). *McGraw-Hill series in psychology. Psychometric theory.* New York, NY, US: McGraw-Hill.
Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436.
Paulson Gjerde, K., Padgett, M. Y., & Skinner, D. (2017). The impact of process vs. Outcome feedback on student performance and perceptions. *Journal of Learning in Higher Education, 13*(1), 73–82.
Pintrich, P. R. (2000). *The role of goal orientation in self-regulated learning*.
Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory and Practice, 41*(4), 219. https://doi.org/10.1207/s15430421tip4104_3.
Popelka, E. (2015). *Improving the accuracy of middle school students' self-assessment, peer assessment, and mathematics achievement*.
Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45*(3), 269–286. https://doi.org/10.1080/00313830120074206.
Rashid, T., & Asghar, H. M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior, 63*, 604–612. https://doi.org/10.1016/j.chb.2016.05.084.
Ridder, J., McGaghie, W. C., Stokking, K. M., & Cate, O. T. (2015). Variables that affect the process and outcome of feedback, relevant for medical training: A meta-review. *Medical Education, 49*(7), 658–673.
Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333–341.
Rienties, B., Toetenel, L., & Bryan, A. (2015). Scaling up learning design: Impact of learning design activities on LMS behavior and performance. *Paper presented at the proceedings of the fifth international conference on learning analytics and knowledge*.
Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning, 4*(1), 33–45.
Sedrakyan, G., & Snoeck, M. (2016). Cognitive feedback and behavioral feedforward automation perspectives for modeling and validation in a learning context. *Paper presented at the international conference on model-driven engineering and software development*.
Segedy, J. R. (2014). *Adaptive scaffolds in open-ended computer-based learning environments.* Vanderbilt University.
Siadaty, M., Gašević, D., & Hatala, M. (2016). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics, 3*(1).
Slingerland, M., Weeldenburg, G., Raijmakers, B., Borghouts, L., & Vos, S. (2016). Authentic learning tasks within physical education teacher education: Redesigning the curriculum. *Paper presented at the 2016 AIESEP international conference, 8-11 June 2016, Wyoming, US*.
Stolp, S., & Zabrucky, K. M. (2017). Contributions of metacognitive and self-regulated learning theories to investigations of calibration of comprehension. *International Electronic Journal of Environmental Education, 2*(1), 7–31.
Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*(4), 437–475.
Studer, M., Mueller, N. S., Ritschard, G., & Gabadinho, A. (2010). Classer, discriminer et visualiser des séquences d'événements. *Paper presented at the EGC*.
Szabo, C., Falkner, N., Knutas, A., & Dorodchi, M. (2018). Understanding the effects of lecturer intervention on computer science student behaviour. *Paper presented at the proceedings of the 2017 ITiCSE conference on working group reports*.
Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167.
Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66.
Timmers, C. F., Walraven, A., & Veldkamp, B. P. (2015). The effect of regulation feedback in a computer-based formative assessment on information problem solving. *Computers & Education, 87*, 1–9.
Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265–283.
Van Laer, S., & Elen, J. (2016). *Adults' self-regulatory behaviour profiles in blended learning environments and their implications for design.* Technology: Knowledge and Learning.
Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475–511.
Veenman, M. V. (1993). *Intellectual ability and metacognitive skill: Determinants of discovery learning in computerized learning environments: Faculteit psychologie.* Universiteit van Amsterdam.
Veenman, M. V., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences, 29*, 123–130. https://doi.org/10.1016/j.lindif.2013.01.003.
Veenman, M. V., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*(2), 187–209. https://doi.org/10.1016/S0959-4752(96)00025-4.
Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and*

Instruction, 14(1), 89–109.
Verpoorten, D., Westera, W., & Specht, M. (2017). Effects of isolated versus combined learning enactments in an online course. *International Journal of Technology Enhanced Learning, 9*(2–3), 169–185.
Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice, 93*, 27–30.
Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551–572.
Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Paper presented at the Proceedings of the third international conference on learning analytics and knowledge*.
Wolters, C. A., Won, S., & Hussain, M. (2017). Examining the relations of time management and procrastination within a model of self-regulated learning. *Metacognition and learning, 12*(3), 381–399.
Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied, 23*(3), 263.
Zhou, M. (2016). *Data pre-processing of student e-learning logs Information Science and Applications (ICISA) 2016.* Springer.
Zimmerman, B. J. (2013a). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147. https://doi.org/10.1080/00461520.2013.794676.
Zimmerman, B. J. (2013b). *Theories of self-regulated learning and academic achievement: An overview and analysis Self-regulated learning and academic achievement.* Routledge10–45.
Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives.* Routledge.
Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2015). A personal agency view of self-regulated learning. *Self-Concept, Motivation and Identity: Underpinning Success with Research and Practice, 83*.